

## 网络附录：统计（马氏）距离

为了更好地理解  $F$  检验的置信域，本附录对统计（马氏）距离及其等高线进行较为详细的介绍，想进一步深入学习相关知识的读者可以参考 [Johnson et al. \(2007\)](#)。

为了便于可视化，我们先考虑二维的情况。通常，在直角坐标系中点  $P = (x_1, x_2)$  到原点  $O = (0, 0)$  之间的距离定义为：

$$d(O, P) = \sqrt{x_1^2 + x_2^2},$$

即， $P$  和  $O$  两点之间的欧氏距离（Euclidean distance）。在欧氏距离的计算中，各个坐标到原点的距离是同等对待的。例如， $(2, 3)$  和  $(3, 2)$  到原点的距离是相同的。但是，若  $x_1$  和  $x_2$  是回归模型中的两个解释变量，其观测数据具有不同的量纲和波动情况，计算距离时同等对待  $x_1$  和  $x_2$  是有问题的。考虑散点图 2.8，其中， $x_1$  样本数据的波动大于  $x_2$  样本数据的波动。我们在图中选取点  $P = (0.5, 0.5)$ ，从欧氏距离的角度讲，点  $P$  在两个坐标轴上的投影长度相等。但是，在  $x_1$  轴方向距离原点 0.5 发生的概率要远大于在  $x_2$  轴方向距离原点 0.5 发生的概率。因此，在计算距离时，我们应该赋予  $x_2$  轴更大的权重。因为数据波动小的方向要赋予更大的权重，所以，一个直观的赋权方式是以标准差的倒数作为权重。于是，一种统计意义上的  $P$  到  $O$  之间的距离可以定义为：

$$d(O, P) = \sqrt{\left(\frac{x_1}{\sqrt{s_{11}}}\right)^2 + \left(\frac{x_2}{\sqrt{s_{22}}}\right)^2}, \quad (2.42)$$

其中， $s_{11}$  和  $s_{22}$  分别是  $x_1$  和  $x_2$  观测数据的样本方差。上式定义的距离实际上是马氏距离的特例，其等高线（到原点距离等于常数  $c$  的点构成的曲线）为：

$$\left\{ (x_1, x_2) : \sqrt{\left(\frac{x_1}{\sqrt{s_{11}}}\right)^2 + \left(\frac{x_2}{\sqrt{s_{22}}}\right)^2} = c \right\},$$

即，

$$\left\{ (x_1, x_2) : \frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}} = c^2 \right\}.$$

如图 2.9 所示，在直角坐标系中，此等高线是中心为原点的椭圆，其长半轴方向与横轴一致，长度为  $c\sqrt{s_{11}}$ ；短半轴方向与纵轴一致，长度为  $c\sqrt{s_{22}}$ 。

式 (2.42) 定义的二维统计距离仅对相互独立的  $x_1$  和  $x_2$  适用，它可以很容易地推广到各个坐标变量相互独立的多维情形。令  $P = (x_1, x_2, \dots, x_p)$  和  $Q = (y_1, y_2, \dots, y_p)$  为  $p$  维欧氏空间中的点， $Q$  为固定点， $P$  可以看成是随机向量  $(x_1, x_2, \dots, x_p)$  的一次实现， $x_1, x_2, \dots, x_p$  相互独立。令  $s_{11}, s_{22}, \dots, s_{pp}$  是由  $(x_1, x_2, \dots, x_p)$  的  $n$  个样本计算出的各个坐标的样本方差，则  $P$  到  $Q$  之间的统计距离为：

$$d(P, Q) = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + \dots + \frac{(x_p - y_p)^2}{s_{pp}}}$$

所有与  $Q$  的距离等于常数  $c$  的点  $P$  构成了一个超椭球（hyperellipsoid），中心为  $Q$ ，各个半轴与坐标轴平行。

若  $x_1$  和  $x_2$  不独立，式 (2.42) 将不再适用。但是，如图 2.10 所示，我们可以将原坐标系旋转到  $(\tilde{x}_1, \tilde{x}_2)$  并计算点  $P$  及所有样本点在新坐标系下的坐标，点  $P$  到原点的距离可

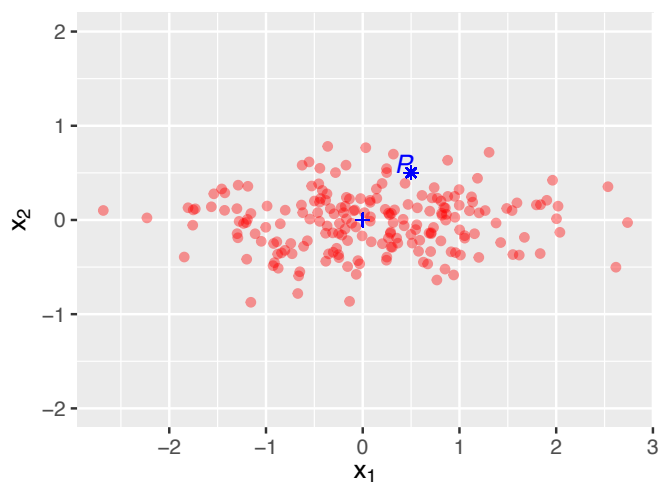


图 2.8:  $x_1$  和  $x_2$  的数据散点图 ( $x_1$  和  $x_2$  相互独立)

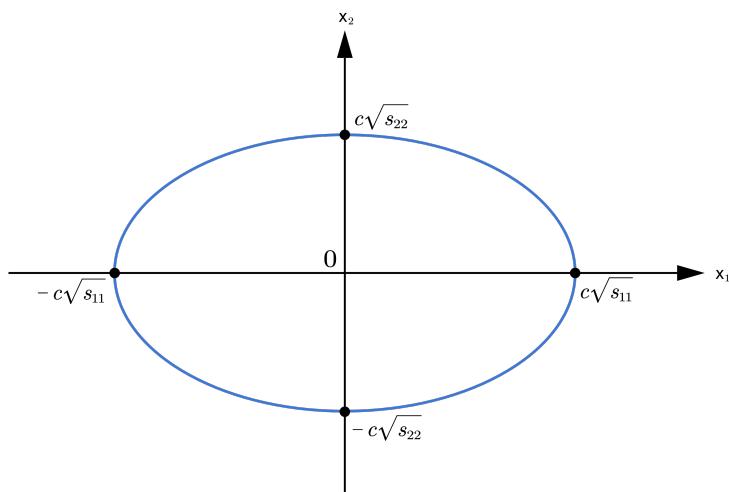


图 2.9: 满足  $\frac{x_1^2}{s_{11}} + \frac{x_2^2}{s_{22}} = c^2$  的等高线

以表示为:

$$d(O, P) = \sqrt{\frac{\tilde{x}_1^2}{\tilde{s}_{11}} + \frac{\tilde{x}_2^2}{\tilde{s}_{22}}}, \quad (2.43)$$

其中,  $(\tilde{x}_1, \tilde{x}_2)$  为  $P$  在新坐标系下的坐标,  $\tilde{s}_{11}$  和  $\tilde{s}_{22}$  是由样本点在新坐标系下的坐标计算的样本方差。若坐标系旋转的角度为  $\theta$ , 则原坐标系下的点  $(x_1, x_2)$  在新坐标系下的坐标  $(\tilde{x}_1, \tilde{x}_2)$  可以表示为:<sup>15</sup>

$$\begin{aligned} \tilde{x}_1 &= x_1 \cos(\theta) + x_2 \sin(\theta) \\ \tilde{x}_2 &= -x_1 \sin(\theta) + x_2 \cos(\theta) \end{aligned}$$

根据此坐标变换公式, 式 (2.43) 给出的距离可以利用原坐标系下的坐标进行表示:

$$d(O, P) = \sqrt{a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2}, \quad (2.44)$$

其中,  $a_{11}, a_{12}$  和  $a_{22}$  由  $\theta, s_{11}, s_{22}$  以及按原坐标系下的坐标计算的样本协方差  $s_{12}$  唯一确定<sup>16</sup>。注意到, 若我们定义对称矩阵  $\mathbf{A}$  为:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix},$$

则式 (2.44) 可以表示为:

$$d(O, P) = \sqrt{\mathbf{x}'\mathbf{A}\mathbf{x}}, \quad \mathbf{x} = (x_1, x_2)'$$

按照式 (2.43) 中距离的定义, 到原点距离等于常数  $c$  的点构成的曲线也是一个椭圆, 它以原点为中心, 半轴方向与  $\tilde{x}_1$  和  $\tilde{x}_2$  的方向一致。

一般地, 令  $P = (x_1, x_2, \dots, x_p)$  和  $Q = (y_1, y_2, \dots, y_p)$  为  $p$  维欧氏空间中的点,  $Q$  为固定点,  $x_1, x_2, \dots, x_p$  不相互独立,  $P$  到  $Q$  之间的统计距离为:

$$d(P, Q) = \sqrt{(\mathbf{x} - \mathbf{y})'\mathbf{A}(\mathbf{x} - \mathbf{y})}, \quad \mathbf{x} - \mathbf{y} = (x_1 - y_1, x_2 - y_2, \dots, x_p - y_p)',$$

其中,

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{12} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{pp} \end{pmatrix}.$$

显然,  $\mathbf{A}$  是对称矩阵。同时, 为了满足  $\mathbf{x} \neq \mathbf{y}$  时  $d(P, Q)$  大于 0, 我们要求  $\mathbf{A}$  是正定的。此距离对应的等高线是  $p$  维欧氏空间中以  $Q$  为中心的超椭圆。接下来我们分析此超椭圆的半轴方向及长度。

为了便于理解, 我们先假设  $\mathbf{y} = \mathbf{0}$ , 同时令  $p = 2$ 。因为  $\mathbf{A}$  是对称阵, 由矩阵的谱分解定理, 我们有:

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}' = \lambda_1\mathbf{e}_1\mathbf{e}_1' + \lambda_2\mathbf{e}_2\mathbf{e}_2', \quad (2.45)$$

其中,  $\mathbf{U} = (\mathbf{e}_1, \mathbf{e}_2)$  是正交矩阵, 由彼此正交的单位特征向量  $\mathbf{e}_1$  和  $\mathbf{e}_2$  构成;  $\mathbf{\Lambda} =$

<sup>15</sup> 参考丘维声 (2015) 中的第四章。

<sup>16</sup>  $a_{11}, a_{12}$  和  $a_{22}$  的具体表达式见 Johnson et al. (2007) 的第 35 页。

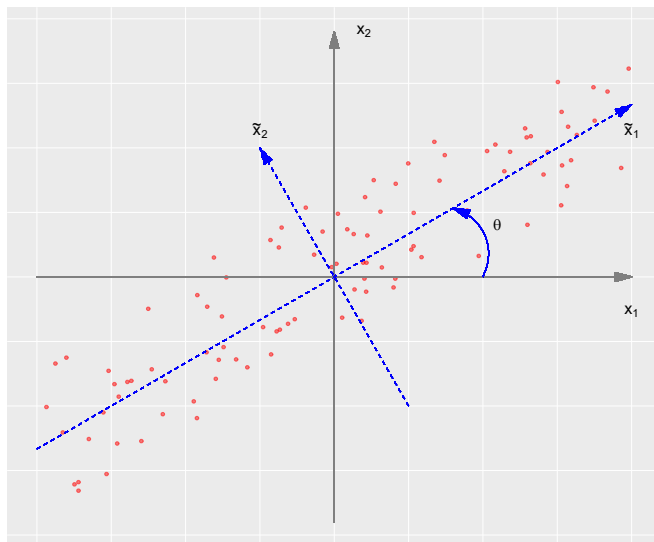


图 2.10:  $x_1$  和  $x_2$  的数据散点图 ( $x_1$  和  $x_2$  不独立)

$\text{diag}(\lambda_1, \lambda_2)$  是对角阵, 对角线上的元素  $\lambda_1$  和  $\lambda_2$  分别是  $\mathbf{e}_1$  和  $\mathbf{e}_2$  对应的特征值, 因为  $\mathbf{A}$  正定, 所以,  $\lambda_1$  和  $\lambda_2$  都大于 0, 不妨设  $1 \leq \lambda_1 < \lambda_2$ 。于是,

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \lambda_1(\mathbf{x}'\mathbf{e}_1)^2 + \lambda_2(\mathbf{x}'\mathbf{e}_2)^2 \equiv \frac{\tilde{x}_1^2}{1/\lambda_1} + \frac{\tilde{x}_2^2}{1/\lambda_2}, \quad (2.46)$$

其中,  $\tilde{x}_1 = \mathbf{x}'\mathbf{e}_1$ ,  $\tilde{x}_2 = \mathbf{x}'\mathbf{e}_2$ 。因此, 满足  $\mathbf{x}'\mathbf{A}\mathbf{x} = c^2$  的等高线为椭圆, 其长半轴长度为  $c\lambda_1^{-1/2}$ ; 短半轴长度为  $c\lambda_2^{-1/2}$ 。注意到  $\mathbf{x} = c\lambda_1^{-1/2}\mathbf{e}_1$  满足  $\mathbf{x}'\mathbf{A}\mathbf{x} = c^2$ , 其欧氏距离长度为  $\sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{c^2\lambda_1^{-1}\mathbf{e}_1'\mathbf{e}_1} = c\lambda_1^{-1/2}$ , 因此,  $\mathbf{e}_1$  的方向即为椭圆的长半轴方向; 同理,  $\mathbf{e}_2$  的方向即为椭圆的短半轴方向。

类似地, 如果  $p > 2$ , 与原点  $O$  的统计距离为常数  $c = \sqrt{\mathbf{x}'\mathbf{A}\mathbf{x}}$  的  $\mathbf{x}$  构成的等高线为中心是原点的超椭球, 它的各个半轴方向与矩阵  $\mathbf{A}$  的特征向量的方向一致, 在  $\mathbf{e}_i$  方向的半轴长度为  $c\lambda_i^{-1/2}$ , 其中,  $\lambda_i$  是  $\mathbf{e}_i$  对应的特征值。更一般地, 若  $Q \neq O$ , 与  $Q$  的统计距离为常数  $c$  的点  $\mathbf{x}$  构成的等高线为中心是  $Q$  的超椭球, 它的各个半轴与矩阵  $\mathbf{A}$  的特征向量平行, 平行于  $\mathbf{e}_i$  的半轴长度为  $c\lambda_i^{-1/2}$ 。

由前文的分析可知统计距离与各个坐标的观测数据的样本方差和样本协方差有关, 事实上, 若记一组样本  $(x_{i1}, x_{i2}, \dots, x_{ip})_{i=1}^n$  的样本协方差矩阵为  $\mathbf{S}$ , 则  $\mathbf{A}$  等于  $\mathbf{S}^{-1}$ 。样本  $\mathbf{x}$  到固定点  $\mathbf{y}$  的马氏距离定义为  $\sqrt{(\mathbf{x} - \mathbf{y})'\mathbf{S}^{-1}(\mathbf{x} - \mathbf{y})}$ 。

此外, 由式 (2.43), 在各个坐标变量不相互独立的情况下, 马氏距离的定义涉及到坐标变换, 并且, 其等高线的半轴长度与数据的波动情况有关。接下来, 我们从坐标变换和数据波动两个角度进一步分析马氏距离及其等高线。

## 坐标变换

我们只详细讨论中心在原点且  $p = 2$  的情况。在式 (2.43) 中,  $d(O, P)$  的计算利用了新坐标系下的坐标, 新坐标系的横轴和纵轴即为距离原点等距的等高线的两个半轴。由谱分解定理, 我们有:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{U}\mathbf{\Lambda}\mathbf{U}'\mathbf{x}.$$

注意到向量  $\mathbf{x}$  左乘矩阵  $\mathbf{U}'$  的几何意义是对向量  $\mathbf{x}$  进行旋转和缩放（见 Calafiore et al. (2014) 第 64 页），我们可以通过研究向量  $\mathbf{e}_x = (1, 0)'$  和  $\mathbf{e}_y = (0, 1)'$ （原坐标系坐标轴方向的单位向量）左乘矩阵  $\mathbf{U}'$  后得到的向量方向来确定新坐标系坐标轴的方向。我们只说明向量  $\mathbf{e}_x$  的情况， $\mathbf{e}_y$  的分析与此类似。令  $\mathbf{y} \equiv \mathbf{U}'\mathbf{e}_x$ ，因为  $\mathbf{U}$  是正交矩阵，左乘  $\mathbf{U}'$  不改变向量的欧氏长度：

$$\|\mathbf{y}\| = \sqrt{\mathbf{y}'\mathbf{y}} = \sqrt{\mathbf{e}_x'\mathbf{U}\mathbf{U}'\mathbf{e}_x} = 1,$$

其中，最后一个等号成立利用了  $\mathbf{U}\mathbf{U}'$  为单位阵。令  $\mathbf{y}$  和  $\mathbf{e}_x$  夹角为  $\theta$ ，根据向量间夹角的余弦公式，我们有：

$$\cos(\theta) = \frac{\mathbf{y}'\mathbf{e}_x}{\|\mathbf{y}\| \cdot \|\mathbf{e}_x\|} = \mathbf{y}'\mathbf{e}_x = \mathbf{e}_x'\mathbf{U}\mathbf{e}_x = \mathbf{e}_x'(\mathbf{e}_1, \mathbf{e}_2)\mathbf{e}_x = e_{11},$$

其中， $e_{11}$  为特征向量  $\mathbf{e}_1$  的第一个元素。另一方面，特征向量  $\mathbf{e}_1$  与  $\mathbf{e}_x$  的夹角余弦为：

$$\frac{\mathbf{e}_1'\mathbf{e}_x}{\|\mathbf{e}_1\| \cdot \|\mathbf{e}_x\|} = \mathbf{e}_1'\mathbf{e}_x = e_{11}.$$

因此，新坐标系的横轴方向即为  $\mathbf{e}_1$  的方向。同理，新坐标系的纵轴方向即为  $\mathbf{e}_2$  的方向。

此外，由定理 A.12，我们有：

$$\lambda_1 \leq \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \leq \lambda_2,$$

并且，当  $\mathbf{x} = \mathbf{e}_2$  时， $\mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{x}$  取到最大值  $\lambda_2$ ；当  $\mathbf{x} = \mathbf{e}_1$  时， $\mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{x}$  取到最小值  $\lambda_1$ 。因为向量  $\mathbf{e}_1$  满足：

$$\frac{\mathbf{e}_1'\mathbf{A}\mathbf{e}_1}{\mathbf{e}_1'\mathbf{e}_1} = \lambda_1.$$

因此，考虑  $\mathbf{e}_1$  方向上满足  $\mathbf{x}'\mathbf{A}\mathbf{x} = c^2$  的点，若假设原点到此点的向量为  $l_1\mathbf{e}_1$ ，则有：

$$c^2 = (l_1\mathbf{e}_1)'\mathbf{A}(l_1\mathbf{e}_1) = l_1^2\lambda_1\mathbf{e}_1'\mathbf{e}_1 = l_1^2\lambda_1.$$

所以， $l_1 = c\lambda_1^{-1/2}$ 。同理，原点到  $\mathbf{e}_2$  方向上满足  $\mathbf{x}'\mathbf{A}\mathbf{x} = c^2$  的点的向量为  $c\lambda_2^{-1/2}\mathbf{e}_2$ 。因为， $\mathbf{e}_1$  和  $\mathbf{e}_2$  是等高线的两个半轴方向，所以，等高线的两个半轴长度分别为  $c\lambda_1^{-1/2}$  和  $c\lambda_2^{-1/2}$ 。

## 样本主成分分析

令  $\mathbf{x}_1, \dots, \mathbf{x}_n$  为  $p$  维随机向量  $\mathbf{x} = (x_1, x_2, \dots, x_p)'$  的观测样本，其样本协方差为  $\mathbf{S}$ 。考虑  $\mathbf{x}$  的  $p$  个线性组合  $y_i = \boldsymbol{\omega}_i'\mathbf{x}$ ， $i = 1, 2, \dots, p$ 。给定  $\mathbf{x}$  的样本， $y_i$  的样本观测值为  $y_{ij} = \boldsymbol{\omega}_i'\mathbf{x}_j$ ， $j = 1, \dots, n$ ，记  $\mathbf{y}_i = (y_{i1}, \dots, y_{in})$ ； $\mathbf{y}_i$  的样本均值为：

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^n \boldsymbol{\omega}_i'\mathbf{x}_j = \boldsymbol{\omega}_i'\bar{\mathbf{x}}, \quad \bar{\mathbf{x}} = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j;$$

$\mathbf{y}_i$  的样本方差为

$$\begin{aligned} s_{ii} &= \frac{1}{n-1} \sum_{j=1}^n (\boldsymbol{\omega}_i'\mathbf{x}_j - \boldsymbol{\omega}_i'\bar{\mathbf{x}})^2 \\ &= \boldsymbol{\omega}_i' \left[ \frac{1}{n-1} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' \right] \boldsymbol{\omega}_i = \boldsymbol{\omega}_i'\mathbf{S}\boldsymbol{\omega}_i. \end{aligned}$$

同理,  $y_i$  和  $y_k$  的样本协方差为:

$$s_{ik} = \omega_i' S \omega_k.$$

### 定义 2.6. 样本主成分

给定样本矩阵  $\mathbf{X}$ , 样本第一主成分  $y_1 = \omega_1' \mathbf{x}$  是在  $\omega_1' \omega_1 = 1$  的条件下, 使得  $\omega_1' \mathbf{x}_j$  ( $j = 1, 2, \dots, n$ ) 的样本方差  $\omega_1' S \omega_1$  最大的  $\mathbf{x}$  的线性变换; 样本第二主成分  $y_2 = \omega_2' \mathbf{x}$  是在  $\omega_2' \omega_2 = 1$  和  $\omega_2' \mathbf{x}_j$  与  $\omega_1' \mathbf{x}_j$  ( $j = 1, 2, \dots, n$ ) 的样本协方差  $\omega_1' S \omega_2 = 0$  的条件下, 使得  $\omega_2' \mathbf{x}_j$  ( $j = 1, 2, \dots, n$ ) 的样本方差  $\omega_2' S \omega_2$  最大的  $\mathbf{x}$  的线性变换; 一般地, 样本第  $i$  主成分  $y_i = \omega_i' \mathbf{x}$  是在  $\omega_i' \omega_i = 1$  和  $\omega_i' \mathbf{x}_j$  与  $\omega_k' \mathbf{x}_j$  ( $k < i, j = 1, 2, \dots, n$ ) 的样本协方差  $\omega_k' S \omega_i = 0$  的条件下, 使得  $\omega_i' \mathbf{x}_j$  ( $j = 1, 2, \dots, n$ ) 的样本方差  $\omega_i' S \omega_i$  最大的  $\mathbf{x}$  的线性变换。



根据样本主成分的定义, 第一主成分可以通过求解如下约束最优化问题得到:

$$\begin{aligned} \max_{\tilde{\omega}_1} \quad & \tilde{\omega}_1' S \tilde{\omega}_1 \\ \text{s. t.} \quad & \tilde{\omega}_1' \tilde{\omega}_1 = 1. \end{aligned}$$

定义拉格朗日函数:

$$L(\tilde{\omega}_1, \tilde{\lambda}) = \tilde{\omega}_1' S \tilde{\omega}_1 - \tilde{\lambda}(\tilde{\omega}_1' \tilde{\omega}_1 - 1).$$

将拉格朗日函数对  $\tilde{\omega}_1$  求导, 最优解  $\omega_1$  和  $\lambda$  满足导数等于 0, 则有:

$$\left. \frac{\partial L(\tilde{\omega}_1, \tilde{\lambda})}{\partial \tilde{\omega}_1} \right|_{\tilde{\omega}_1 = \omega_1, \tilde{\lambda} = \lambda} = 2S\omega_1 - 2\lambda\omega_1 = \mathbf{0},$$

即,  $S\omega_1 = \lambda\omega_1$ 。因此,  $\lambda$  是矩阵  $S$  的特征值,  $\omega_1$  是  $\lambda$  对应的单位特征向量。由矩阵的谱分解定理, 最优解满足:

$$\omega_1' S \omega_1 = \lambda.$$

令  $\lambda_1^*$  是  $S$  的最大特征值,  $\omega_1$  是  $\lambda_1^*$  对应的单位特征向量。于是,  $\omega_1' \mathbf{x}$  是第一主成分, 其样本方差等于:

$$s_{11} = \omega_1' S \omega_1 = \lambda_1^*.$$

类似地, 可以证明如下定理:

### 定理 2.15. 样本主成分定理

设  $\mathbf{x}_1, \dots, \mathbf{x}_n$  是  $p$  维随机向量  $\mathbf{x}$  的一组观测, 样本协方差矩阵为  $S$ ,  $S$  的  $p$  个特征值按从大到小排序为  $\lambda_1^* \geq \lambda_2^* \geq \dots \geq \lambda_p^* \geq 0$ , 特征值对应的单位特征向量分别为  $\omega_1, \dots, \omega_p$ , 则样本的第  $k$  主成分 ( $k = 1, \dots, p$ ) 为:

$$y_k = \omega_k' \mathbf{x},$$

其样本方差为  $\lambda_k^*$ 。



假设  $p = 2$ , 我们考虑满足  $\mathbf{x}' A \mathbf{x} = c^2$  的等高线。因为  $A = S^{-1}$ , 若假设  $\lambda_1$  和  $\lambda_2$  是  $A$  的特征值并且满足  $1 \leq \lambda_1 < \lambda_2$ , 则  $S$  的特征值为  $\lambda_1^* = 1/\lambda_1$  和  $\lambda_2^* = 1/\lambda_2$  且满足  $\lambda_1^* > \lambda_2^*$ 。由定理 2.15,  $\lambda_1^*$  和  $\lambda_2^*$  分别是第一主成分和第二主成分的样本方差。由式 (2.46),

在等高线的长半轴方向数据方差最大。